Context

Invasive Ductal Carcinoma (IDC) is the most common subtype of all breast cancers. To assign an aggressiveness grade to a whole mount sample, pathologists typically focus on the regions which contain the IDC. As a result, one of the common pre-processing steps for automatic aggressiveness grading is to delineate the exact regions of IDC inside of a whole mount slide.

Content

The original dataset consisted of 162 whole mount slide images of Breast Cancer (BCa) specimens scanned at 40x. From that, 277,524 patches of size 50 x 50 were extracted (198,738 IDC negative and 78,786 IDC positive). Each patch’s file name is of the format: u\_xX\_yY\_classC.png — > example 10253\_idx5\_x1351\_y1101\_class0.png . Where u is the patient ID (10253\_idx5), X is the x-coordinate of where this patch was cropped from, Y is the y-coordinate of where this patch was cropped from, and C indicates the class where 0 is non-IDC and 1 is IDC.

Acknowledgements

The original files are located here: <http://gleason.case.edu/webdata/jpi-dl-tutorial/IDC_regular_ps50_idx5.zip> Citation: <https://www.ncbi.nlm.nih.gov/pubmed/27563488> and <http://spie.org/Publications/Proceedings/Paper/10.1117/12.2043872>

Inspiration

Breast cancer is the most common form of cancer in women, and invasive ductal carcinoma (IDC) is the most common form of breast cancer. Accurately identifying and categorizing breast cancer subtypes is an important clinical task, and automated methods can be used to save time and reduce error.